An Internship Report

on

PROCESS MINING VIRTUAL INTERSHIP

Submitted in partial fulfilment of the requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering (Data Science)

by

A Pavan Kumar **214G1A3272**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(DATA SCIENCE)

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)

(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.

2023 – 2024

SRINIVASA RAMANUJAN INSTITUTE OF TECHNOLOGY (AUTONOMOUS)

(Affiliated to JNTUA, accredited by NAAC with 'A' Grade, Approved by AICTE, New Delhi & Accredited by NBA (EEE, ECE & CSE))
Rotarypuram village, B K Samudram Mandal, Ananthapuramu-515701.

Department of Computer Science & Engineering (Data Science)

Certificate

This is to certify that the internship report entitled Process Mining Virtual Internship is the bonafide work carried out by **A Pavan Kumar** bearing Roll Number 214G1A3272 in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering (Data Science)** from June 2022 to September 2022.

Internship Coordinator

Mr. P. Veera Prakash, M. Tech., (Ph.D.),
Assistant Professor & HOD of CSE

Head of the Department

Dr. P. Chitralingappa, M. Tech., Ph.D.,
Associate Professor

Date:30-08-23

Place: Ananthapuramu

EXTERNAL EXAMINER

PREFACE

Brief overview of the company's history: All India Council for Technical Education (AICTE) has initiated various activities for promoting industrial internship at the graduate level in technical institutes and Eduskills is a Non-profit organization which enables Industry 4.0 ready digital workforce in India. The vision of the organization is to fill the gap between Academic and Industry by ensuring world class curriculum access to the faculties and students. Formation of the All-India Council for Technical Education (AICTE) in 1945 by the Government of India.

Purpose: With a vision to create an industry-ready workforce who will eventually become leaders in emerging technologies, EduSkills & AICTE launches 'Virtual Internship' program on Process Mining. This field is one of the most in-demand, and this internship will serve as a primer

Company's Mission Statement: The main mission of these initiatives is enhancement of the employability skills of the students passing out from Technical Institutions.

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that I have now the opportunity to express my gratitude for all of them.

It is with immense pleasure that I would like to express my indebted gratitude to my internship coordinator **Mr. P. Veera Prakash, Assistant Professor & HOD, Department of Computer Science and Engineering**, who has supported me a lot and encouraged me in every step of the internship work. I thank him for the stimulating support, constant encouragement and constructive criticism which have made possible to bring out this internship work.

I am very much thankful to **Dr. P. Chitralingappa, Associate Professor & HOD, Computer Science and Engineering (Data Science),** for his kind support and for providing necessary facilities to carry out the work.

I wish to convey my special thanks to **Dr. G. Balakrishna, Principal** of **Srinivasa Ramanujan Institute of Technology** for giving the required information in doing my internship. Not to forget, I thank all other faculty and non-teaching staff, and my friends who had directly or indirectly helped and supported me in completing my internship in time.

I also express our sincere thanks to the Management for providing excellent facilities and support**.**

Finally, I wish to convey my gratitude to my family who fostered all the requirements and facilities that I need.

A. Pavan Kumar

2141G1A3272

INDEX

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
API	Application programming interface
BPM	Business Process Management
CRM	Customer Relationship Management
EMS	Execution Management system
ERP	Enterprise Resource Planning
GUI	Graphical User Interface
IT	Information Technology
KPI	Key Performance Indicators
PQL	Process Query Language
PDM	Process Diagnostics Method
PMPM	Process Mining Project Methodology
SCM	Supply chain management

CHAPTER - 1

INTRODUCTION

1.1 What is process mining?

Process mining applies data science to discover, validate and improve workflows. By combining data mining and process analytics, organizations can mine log data from their information systems to understand the performance of their processes, revealing bottlenecks and other areas of improvement. Process mining leverages a data-driven approach to process optimization, allowing managers to remain objective in their decision-making around resource allocation for existing processes. Process mining can also be described as a part of business process management (BPM) that applies data science (with its data mining and machine learning techniques) to dig into the records of the company's software, get the understanding of its process's performance, and support optimization activities.

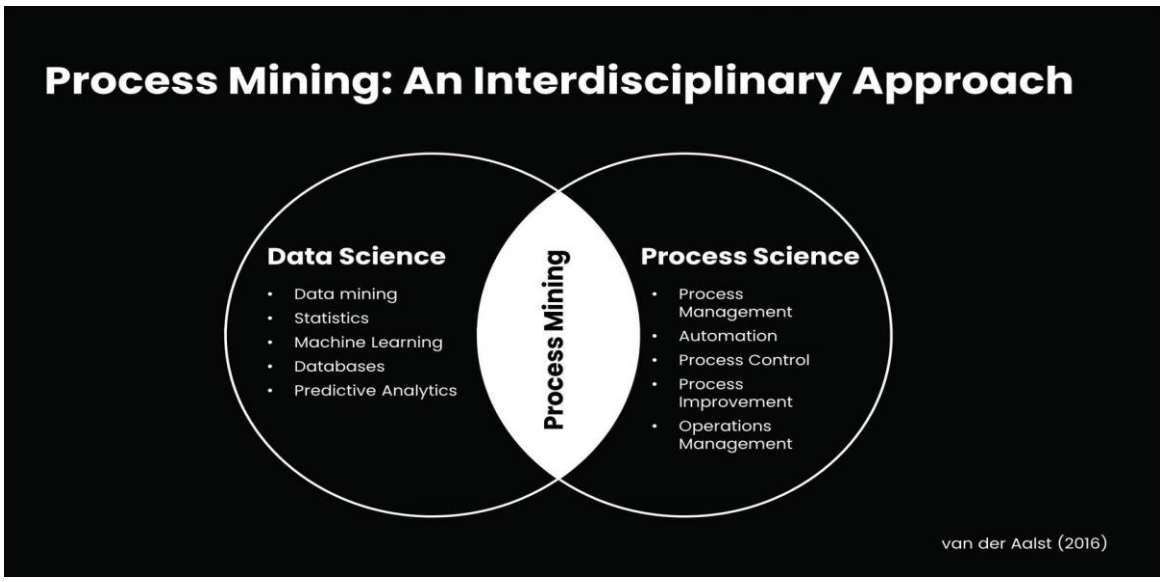

Fig. No. 1.1: Process Mining Approach

1.2 What is process?

One easy example for that could be a pizza delivery process. It starts with placing the order by calling the pizza company or via their website. Then, the order is assigned to a pizzamaker. The pizzamaker bake the pizza, the pizza is packaged, a delivery person delivers it to the assigned address and the payment is received. The problem is this is the ideal scenario of pizza delivery process. But in practice, there are so many things that can go wrong on the way there. The pizzamaker might put the wrong ingredients, the delivery person might go to a different address, or the payment fail. Therefore, we can say processes are the engine of every experience. Understanding these processes and optimizing are crucial for successful businesses.

Fig No. 1.2: Process Example

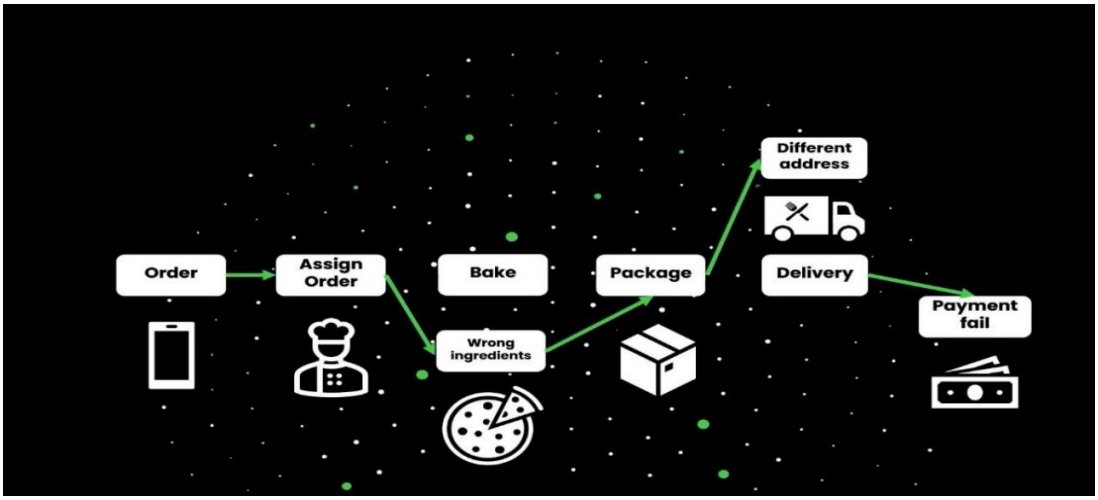

Fig No. 1.3: Process Misleading Phase

1.3 Technical terms in process mining

1. **Process**: Process is very simply a series of linked actions or steps taken in order to achieve a particular end.
2. **Process Model**: Describes a process and is based on the corresponding event data.
3. **Events**: Events are individual occurrences or actions that take place within a process. These actions can be carried out by humans, systems, or machines. Each event represents a specific activity or interaction.
4. **Event data**: Event data in process mining refers to the collection of detailed information about the various events, actions, and interactions that occur within a business process. This data includes timestamps, identifiers, and attributes associated with each event, and it serves as the basis for conducting process mining analyses
5. **Event log**: An event log is a chronologically ordered list of the recorded events. Event logs are defined using Activity tables that contain the Activity, Case and Timestamp data. An Event Log contains each of the three key pieces of information.
    1. Case ID: a unique identifier such as a purchase order item, invoice number or order number.
    2. Activity: the description of what has happened - for example, the creation of a purchase order or the receipt of goods.
    3. Timestamp: the date and time that the activity took place.

CHAPTER – 2

METHODOLOGY

2.1 Introduction

Process mining consists of several distinct phases that collectively enable organizations to analyze, optimize, and improve their business processes. These phases are designed to uncover insights, identify bottlenecks, and facilitate data-driven decision-making. When applied skillfully, the Process Mining methodology, may allow you to map processes, find deviations from usual proceedings, and to discover bottlenecks and opportunities for improving your processes. There is still plenty of chances to refine Process Mining and to employ it successfully, especially in such process demanding areas like supply chain.

2.2 Business Process Management

Business Process Management (BPM) and process mining are closely related concepts that work together to enhance organizational efficiency and effectiveness. BPM focuses on designing, modeling, implementing, and continuously improving business processes, while process mining provides data-driven insights to optimize those processes**.** It is the discipline in which use various methods to discover, various methods to discover, model, analyze, measure, improve, optimize, business Any combination of methods used to manage a company's business processes is BPM. processes can be structured and repeatable or unstructured and variable.

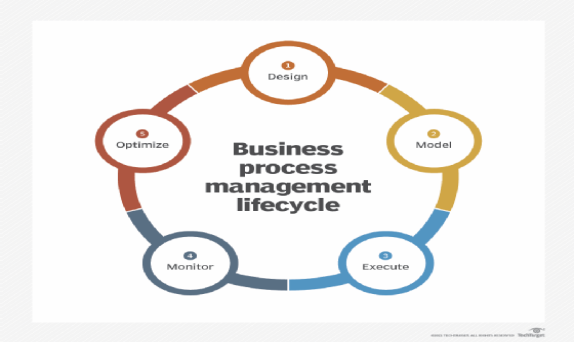

Fig No. 2.1: Life cycle of BPM

2.3 Types of Methodology

There are two well-known Process Mining methodologies: Process Diagnostics Method [PDM], adapted for healthcare environments, useful for providing a fast and broad process overview and the L* life-cycle model, which is more complex than PDM, providing more functionalities, like process improvement and operational support. There is other methodology called PM^2.

2.3.1 Process Diagnostics Method [PDM]

The methodology aims at giving a broad overview of the process in the information system within a short period of time. It consists of six phases: (1) log preparation, in which the event log of the information system is extracted, (2) log inspection, to get a first glance of the process, (3) control flow analysis, (4) performance analysis and (5) role analysis, i.e. the persons and resources that execute activities in the process. (6) Transfer Results. Finally, the results are reported to the client. Figure 1 shows the methodology as a process. The first two stages are input for both control flow analysis and role analysis, performance analysis also needs the input of control flow analysis.

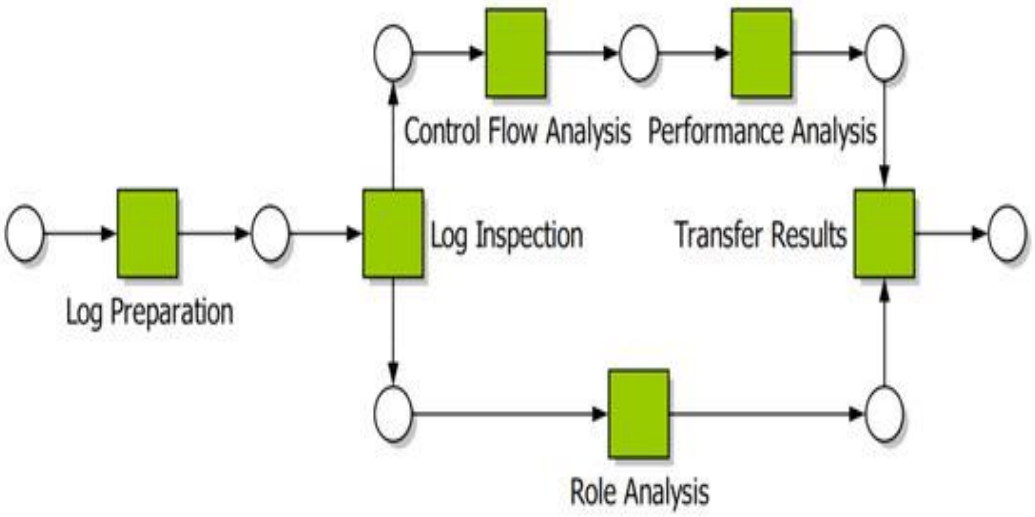

Fig No 2.2: Phases of the PDM methodology

2.3.2 Six phases of PDM

1. **Log preparation** in which the event log is extracted from the organisation's information systems. The information system's internal format of logs can anything from plain text files to internal databases, like in an SAP system. Therefore, preprocessing of the log is needed. Preprocessing already raises many questions. The first step is to select the best notion of a case, since often there are several candidates. The next step is the identification of activities and their events. If the log has multiple time stamps, the semantics of the time stamps needs to be clear, e.g. is it a time stamp of the start of the event, or of the case. These kinds of problems need to be tackled in order to get a proper event log to proceed to the next phase.

2. **Log inspection** from which a basic understanding of the process is developed. In this phase, statistics about the log are gathered. This includes information about the number of cases and roles, the total number of events, the different events present, the minimal, maximal and average number of events per case, the found start and end events and their occurrences, etc. These statistics give insights in the size of the process and event log, and helps to tune mining algorithms and to evaluate the results obtained in next phases. Based on these statistics, the event log is filtered to remove incomplete cases, i.e. cases that were started before the start of the event log, and cases that were still running at the end of the event log are removed This filtered event log is input for the next phases.

3. **Control flow analysis** which involves either checking that the event log conforms to an existing process description/model or automatically discovering a process model from the log. In this phase, the control flow aspect of the process is analyzed. It gives answers to the question "what does the actual process look like?". First, if the organization has a process description, a conformance check is executed to check whether the process is conform

specification, i.e. that each case in the event log can be replayed in the process definition. If either this is not the case, or a process description does not exist, the control flow needs to be discovered. There are many algorithms available to discover processes. Just running different discovery algorithms often gives some process model, but typically, this results in a spaghetti like process model, since infrequent behavior, like exceptions, are also taken into account. Infrequent behavior means that there is a case whose sequence does not occur too often

4. **Performance analysis** including discovering bottlenecks and calculating processing times. After discovery of the control flow, these process models can be used to analyze the performance of the process. This phase answers questions like "are there any bottlenecks in the process?". First, a dotted chart analysis is used to compare cases and their throughput times. The vertical axis of a dotted chart represents the cases in the event log, the horizontal axis represents time. This already gives helpful insights in the performance of the system. Next, by replaying the log on the process model, bottlenecks and throughput times of individual activities and the process itself are calculated.

5. **Role analysis** which provides information on the division of work within the organisation (as it relates to the process being analysed). First, a role-activity matrix is created. In this matrix, the rows represent the roles, the columns represent each event of the event log. Each cell contains the number of times that role executed that event. Next, for each role a profile is made from this matrix. If roles have similar profiles, they form a group. In this way, the roles in the event log can be partitioned into groups. Next, the roles are analyzed to discover specialists, i.e. roles that only execute a few activities, but very frequently, and generalists, i.e. roles that execute many different activities in the process. One way to discover these types of roles is to generate a role hierarchy, a hierarchy based on the different activities' roles execute. A directed arrow between two roles indicate that the role at the base of the arrow can do at least

the activities of the role at the arrow head. Thirdly, a social network analysis is performed. A social network is a (directed) graph representing relations between roles, based on a property discovered in the event log. This property can be handover of work, i.e. roles that pass work of a single case to one and another, and subcontracting, i.e. a role passes work of a single case to another role, and this role returns the work back to the first role after finishing its activity. Important metrics within social network analysis are the betweenness centrality, i.e. the number of roles a role can reach by only traversing directed edges, the degree of incoming edges, and the degree of outgoing edges

6. **Transfer of results** to process owners in such a way as they understand the outcomes, thus allowing the organisation to implement process changes. The aim of this methodology is to gain insights in the processes of the organization and their support by information systems. The outcome of the diagnostics shows all behavior seen in the system. This behavior often deviates from the intended process, because of unwanted behavior, like sequences or user profiles that are not allowed due to e.g. legislation, but also because of smart shortcuts taken by the users of the system to make their work easier. The performer of the diagnostics is not able to make the distinction between wanted and unwanted behavior. Therefore, it is important to discuss the outcome of the diagnostics with the client, so that the client understands the outcome, and gets a better understanding of the information system. With this knowledge, and with the support of the performer of the diagnostics, the client can refactor its information system to make it more efficient or to speed it up, e.g. by deciding to disallow certain sequences or user profiles, or to support certain shortcuts taken by its users.

2.3.3 L* life-cycle model

The L* model consists of 5 stages:

- Justification and planning
- Extract
- Create control-flow model and connect to event log
- Create integrated process model
- Provide operational support

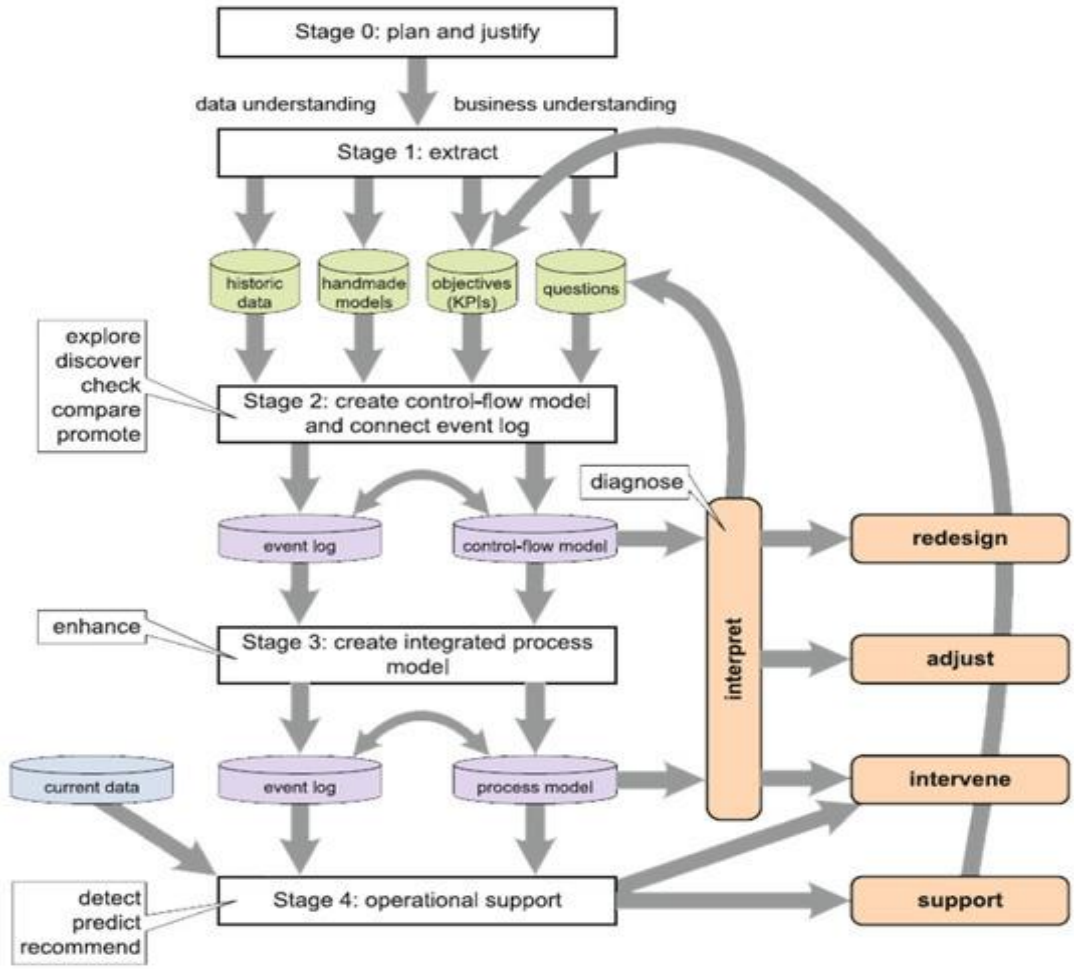

Fig No. 2.3: The L* methodology

2.3.4 Stages of L* methodology

1. **Justification and planning** to clearly outline the reasoning behind the study and to identify resources required for the study.
2. **Extract** domain knowledge (from domain experts and historical data) to (i) develop an understanding of the domain and of the data available for analysis and, (ii) generate artifacts such as hand-made models, objectives and questions.
3. **Create control-flow model and connect to event log** using automated discovery techniques.
4. **Create integrated process model** by extending to other perspectives
5. **Provide operational support** based on insights derived from earlier stages.

2.3.5 PMPM Model or PM^2 Model

PM2 guides organization's performing process mining projects aimed at improving process performance or compliance to rules and regulations. The goals of a process mining project can be very concrete, e.g. achieving a cost reduction of 10% for a given process, or more abstract, e.g. obtaining valuable insights regarding the performance of several processes. Through PM2, these goals are translated into concrete research questions which are iteratively refined and answered, resulting in findings that are the basis of improvement ideas for the selected process.

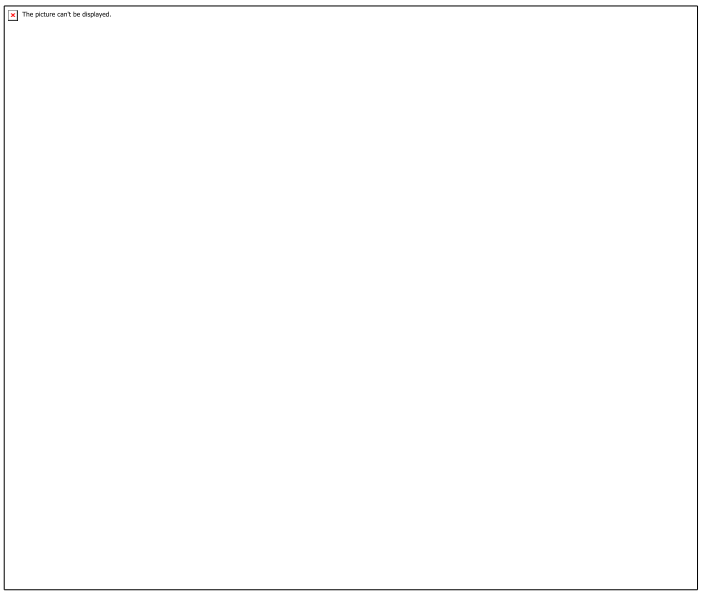

Fig No. 2.4: PMPM Model

2.3.6 Stages of PMPM Model

1. **Planning** Identify the process and gather basic knowledge; Determine the objectives of the project; Determine the required tools and techniques. The objective of the planning stage is to set up the project and to determine the research questions. We consider two main goals for starting process mining projects: improving performance of a business process, or checking its compliance with respect to certain rules and regulations. The inputs of this stage are the organization's business processes. The outputs are goal-related research questions and a set of information systems supporting the execution of the business processes to be analyzed.

2. **Extraction** Locate the required data in the system's logs; Explore the data in the system's logs; Verify the data in the system's logs. The extraction stage aims to extract event data and, optionally, process models. Inputs for this stage are the research questions and the information systems that support the execution of the selected business processes to be analyzed. The outputs of this stage are event data, i.e. a collection of events without predefined case notion or event classes, and possibly process models. Three activities for this stage: determining scope, extracting event data, and transferring process knowledge.

3. **Data Processing** Select the dataset in terms of event context, timeframe and aspects; Extract the set of required data; Prepare the extracted dataset, by cleaning, constructing, merging and formatting the data. The main objective of the data processing stage is to create event logs as different views of the obtained event data and to process event logs in such a way that it is optimal for the mining and analysis stage. In addition to the event data as our main input, one can also use process models as an input to filter the event data. The outputs are event logs that are used in the mining and analysis stage. our main input,

one can also use process models as an input to filter the event data. The outputs are event logs that are used in the mining and analysis stage.

4. **Mining and Analysis** Get familiar with the log by gathering statistics; Make sure that the process contained in the event log is structured enough to apply the required process mining techniques; Apply process mining techniques to answer business questions. Inputs for this stage are event logs. In addition, if process models are available, they can also be used for conformance checking and enhancement activities. Output for this stage are findings that answer research questions related to performance and compliance goals. We identify four types of activity for this stage: process discovery, conformance checking, enhancement and process analytics. The first three activities are well-known process mining techniques. Process analytics are other complementary analysis techniques, e.g. data mining and visual analytics, which can be applied in the context of business processes.

5. **Evaluation** Verify the modelled work; Validate the modelled work; Validate the modelled work; Decide on an elaboration of the process mining project. The objective of the evaluation stage is to relate the analysis findings to improvement ideas that achieve the project's goals. The inputs are the process models, performance and compliance findings from the analysis stage. The outputs are improvement ideas or new research questions. The activities for this stage are: Diagnose, and Verify & Validate (V&V).

6. **Deployment** Identify if and how the process can be improved by improvement actions; Present the project results to the organization. The objective of the process improvement & support stage is to use the gained insights to modify the actual process execution. The inputs of this stage are the improvement ideas from the evaluation stage. The outputs of this stage are process modifications. The activities are: implementing improvements and supporting operations

2.4 Types of Process mining techniques

There are three types of process mining, which are discovery, conformance, and enhancement.

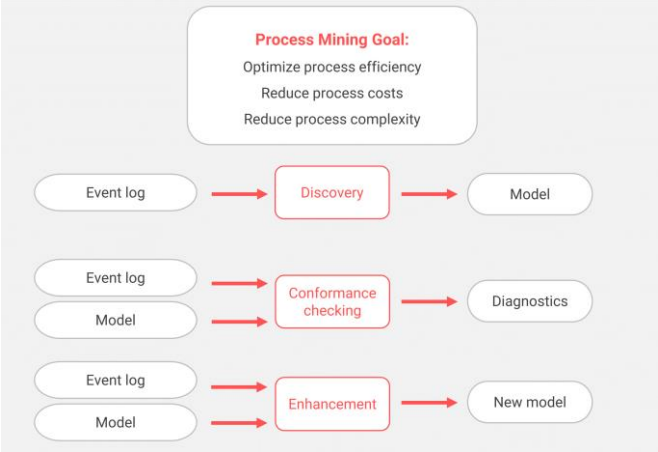

Fig No. 2.5: Process mining types

2.4.1 Process discovery:

Process discovery is a process mining technique used to create a data-based visualization of process workflows. Using data found in event logs, process mining automatically generates a discovered model for analysis, giving users a visual and unbiased representation of their business processes. The primary goal of process discovery is to increase transparency and process knowledge for enhancement.

Working of process discovery: Process discovery needs data to work. These data points include names and descriptions of the activities or events in a process, time stamps for the start and end times of activities, and other process attributes such as case IDs (depending on the available data).

Once identified, data can be exported, transformed, and imported into the process mining system for process discovery. Business process discovery outcomes are then visualized in the discovered model to uncover optimization and process automation opportunities.

Steps in process discovery:

- Checking for availability of data points.

- Extracting and transforming data.
- Importing data into the process mining system.
- Generating a visualization of the discovered process, or process model.

Use cases:

- Financial management
- Risk management
- Compliance management
- Quality assurance and control
- Customer experience and onboarding

2.4.2 Conformance Checking

Conformance checking is a technique used to check process compliance by comparing event logs for a discovered process with the existing reference model (target model) of the same process. This technique is used to determine whether the target process corresponds to the actual process, highlighting deviations between the two.

Purpose: The purpose of conformance checking is to identify two types of discrepancies:

- Unfitting log behaviour: behaviour observed in the log that is not allowed by the model.
- Additional model behaviour: behaviour allowed in the model but never observed in the log.

There are broadly three families of techniques for detecting unfitting log behaviour: replay, trace alignment and behavioural alignment.

Working:

In process mining, business processes are visualized using event logs. This visualization is also referred to as the discovered model. All logged process activities, as well as their durations and sequences, are represented in this process model.

Conformance checking compares the discovered model with a reference or target model, which can be visualized as a superimposition of the two models.

While it's ideal to see a high degree of similarity between the basic activities in both models, this isn't always the case. In reality, process workflows can include extra, unnecessary steps, and steps can be missed or performed in the wrong order. Conformance checking is designed to identify these cases.

2.4.3 Enhancement

Process enhancement, sometimes referred to as model enhancement, is a process mining technique that's used to extend or enhance a target model or reference model using discovered information about the actual process. For example, analysis may uncover bottlenecks or unplanned process sequences that can be eliminated from the model to make it a better representation of the ideal process.

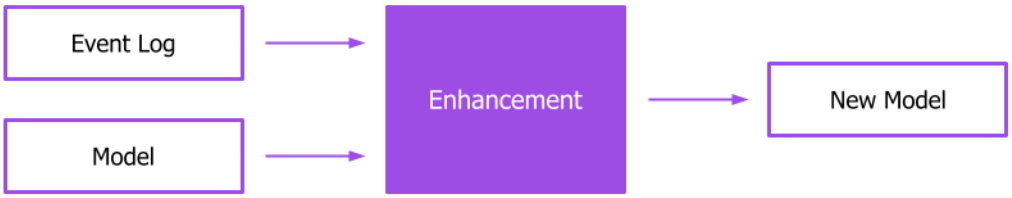

Fig No. 2.6: Process enhancement

Working:

A data-driven process model is created using a log file of the actual process. Based on this information, improvements or changes are made. For example, bottlenecks or unplanned process sequences identified in this way can be eliminated. The objective of process enhancement is to optimize the process model and the process itself.

Steps in Process enhancement:

1. Analyse the process data.
2. Identify optimization potentials (process discovery, conformance checking).
3. Adapt the target process model (process enhancement).

4. Implement the target process model.
5. Check the new process implementation on the basis of the analysis of the process data (continuous improvement process).

CHAPTER 3

MODULES

3.1 Introduction to Process Mining

This module introduces the process mining field and its vast variety of applications. Process Mining is the combination of two disciplines: Data Science and Business Process Management. Process Mining essentially uses Data Science techniques, such as Big Data and AI, to address Process Science problems such as process improvement and automation (cf. van der Aalst 2016). Event Logs are the format in which we can retrieve our digital footprints from the underlying IT systems. They're essentially the log books that IT systems keep to record what events take place for each Case ID and at what time. The Event Log information can be retrieved from several types of IT systems such as Enterprise Resource Planning (ERP), Supply Chain Management (**S**CM) or Customer Relationship Management (CRM) systems. These systems typically generate and store Event Log information in real time. Event Log information might also be retrieved in various situations and contexts from automated payment to customer journeys.

3.2 Foundation of Process Mining

3.2.1 Variant Explorer

Variant is an object which have connections of products or object each other in the portal space. You see the full picture of your entire process. If you think this looks like a bowl of spaghetti, you're not the only one! It can be a bit overwhelming, but at times you might find it useful. You can zoom in on an activity and filter on it. If we think about a process as a road trip, each process variant would be a potential route. Each activity within a process would be a waypoint along a route, and the connections between activities are like the roads that connect the stops. And, each trip a person makes along a particular route would be a case. In short, Variant

Explorer gives you a quick way to see whether most process cases follow an acceptable flow of activities or not and helps you develop your first analysis questions

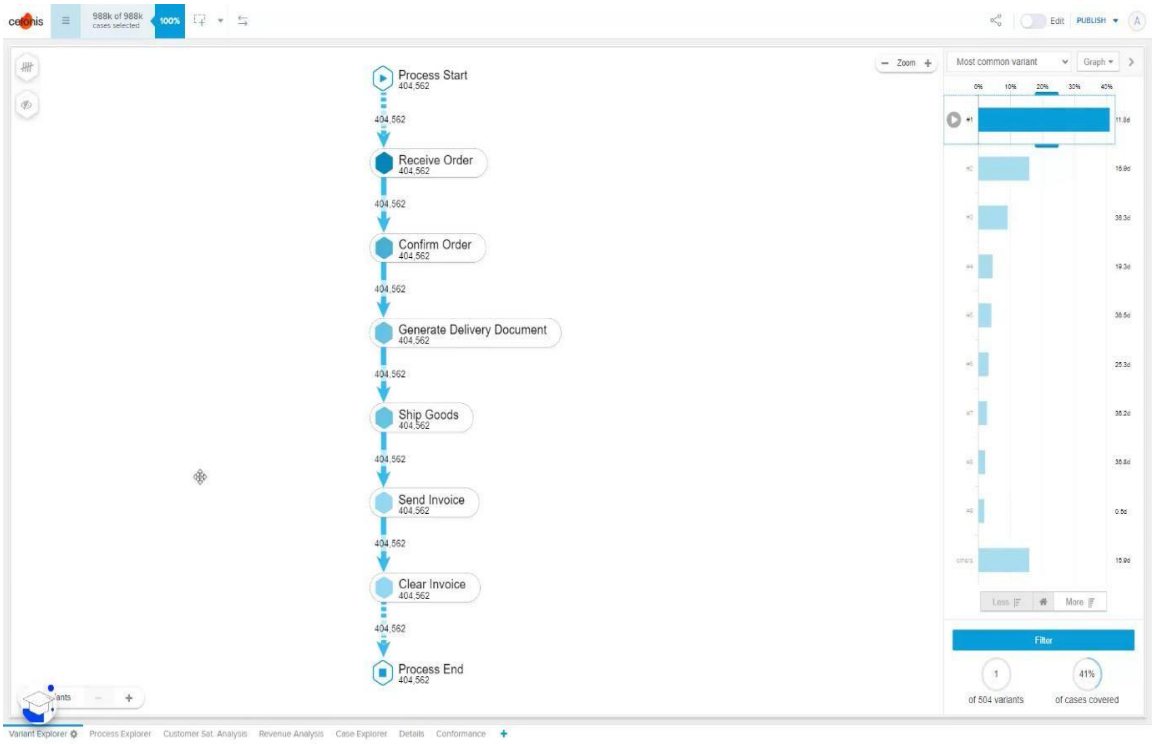

Fig No. 3.1: Variant Explorer

3.2.2 Process Explorer

The Process Explorer is another analysis tool to use when taking an exploratory approach. It's especially useful for quickly revealing activities beyond the most common ones. It also allows you to narrow your focus on a single activity, for example an undesired activity, to see which activities cases typically come from and which activities they're going to. You'll be nicely surprised by all that you can accomplish with it.

In the Process Explorer, if you display the Throughput Time KPI, you are looking at the time it took all the cases in the analysis to go directly between the two displayed

activities. That is unlike in the Variant Explorer where the time is reflective of the cases in the variant or variants selected.

These metrics and KPIs are customizable by the person who creates the analysis. A common custom KPI is automation rate; that is the percentage of time when the activity was completed automatically and not manually

3.2.3 Charts and Tables, Review KPIs

- A **dimension** is a category of attributes; for example, the dimension "customer name" is a category for individual customer names. Other examples of dimensions, depending on the nature of the process, can include vendor name, sales organization, region, and material group.

- **Key Performance Indicators (KPIs)** are used to calculate and add aggregated values; for example, case count, order value, invoice value, throughput time, and automation rate.

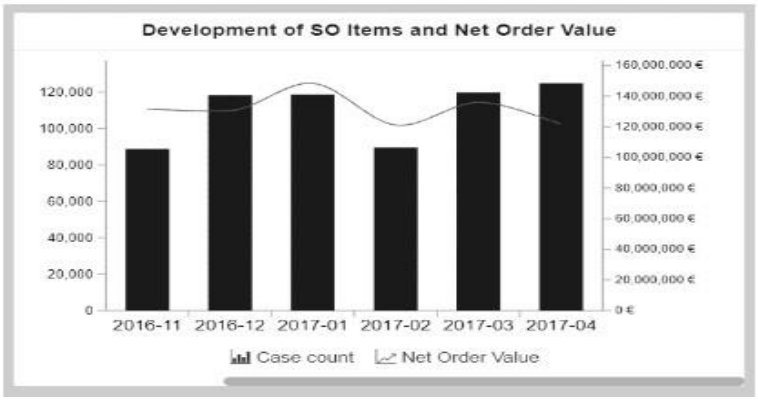

Fig No. 3.2: Items and Cost Graph using KPI'S

3.2.4 Case Explorer

The Case Explorer is useful once you've narrowed down the analysis to a few cases that you want to investigate further. You can view specific case details such as timestamp of activities, user type (manual or automatic), possibly even user name (depending on your setup), and other useful pieces of info.

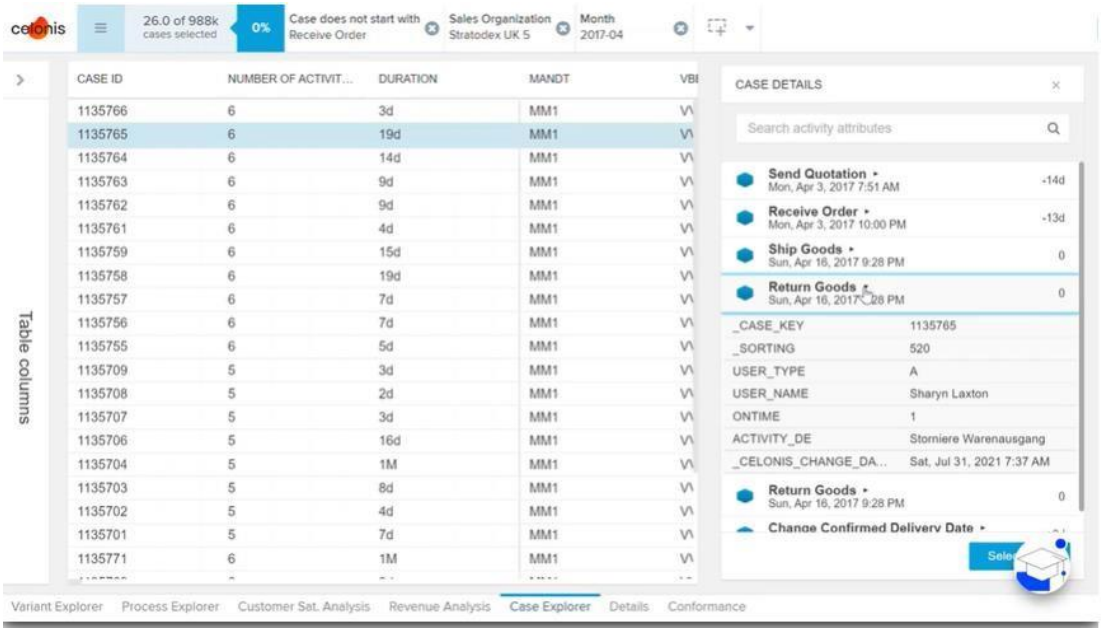

Fig No. 3.3: Case Explorer

3.2.5 Conformance Checker

Every organization has an optimal process in mind that they want to achieve. With the conformance checker, you can see how far away the organization is from reaching that goal and investigate common patterns for inefficiency. The Conformance checker evaluates each case against the process model your organization has specified to determine whether it conforms to it or not. More specifically, the Conformance checker is looking at each case's set and sequence of activities to the one defined in the process model.

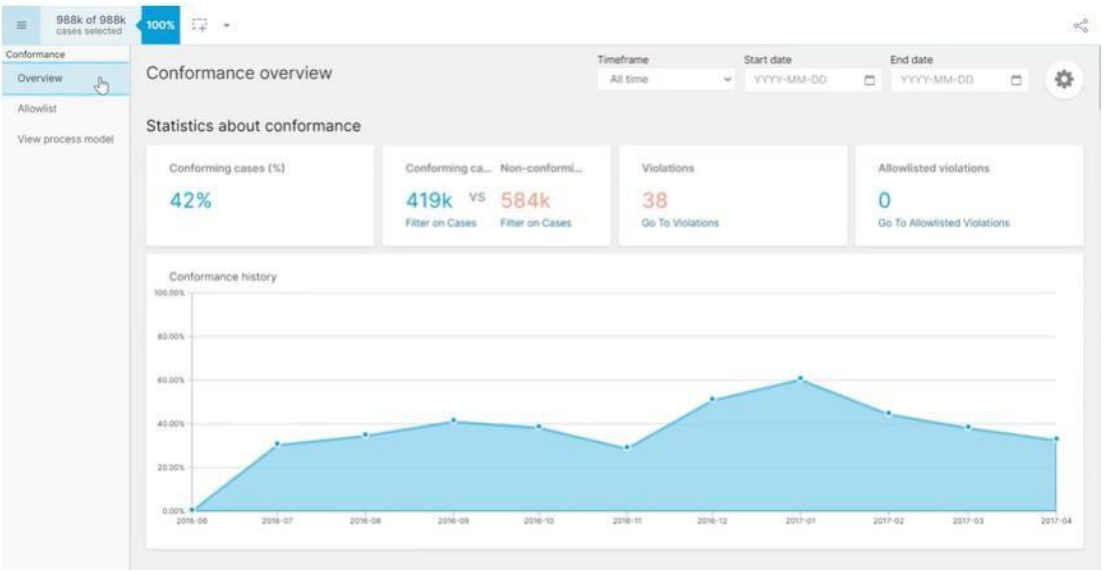

Fig No. 3.4: Conformance Overview

3.3 PQL Queries

The Process Query Language (PQL) is a domain-specific language tailored towards a special process data model and designed for business users. PQL enables the user to translate process-related business questions into queries, which are then executed by a custom-built query engine. PQL covers a broad set of operators, ranging from process-specific functions to aggregations and mathematical operators. Its syntax is inspired by SQL, but specialized for process-related queries

3.3.1 Celonis PQL Engine

Celonis PQL is an integral component of the Celonis Software Architecture. All Celonis applications use this language to query data from a data model. Celonis PQL Engine is an analytical column-store main memory database system. It evaluates Celonis PQL queries over a defined data model.

Applications: Celonis applications provide a variety of tools for the business user to discover, monitor and enhance business processes. All applications use Celonis PQL to query the required data. They include easy-to-use GUIs, providing a convenient way for the users to interact with the process and business data. In the applications, the users can specify custom Celonis PQL queries.

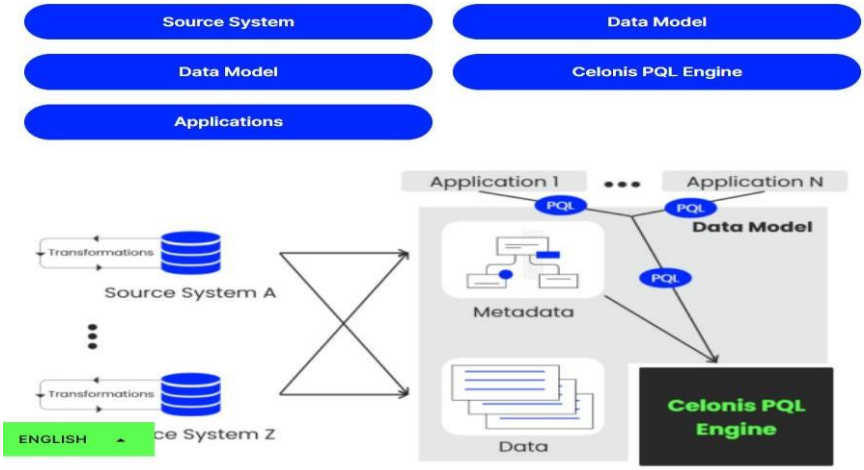

Fig No. 3.5: Celonis PQL Engine Architecture

3.3.2 SQL VS PQL

Even though Celonis PQL is inspired by SQL, there are major differences between the two query languages.

On a high level, Celonis PQL varies along four key dimensions:

1. Language Scope
2. Data Manipulation Language
3. Data Definition language
4. Domain-Specific

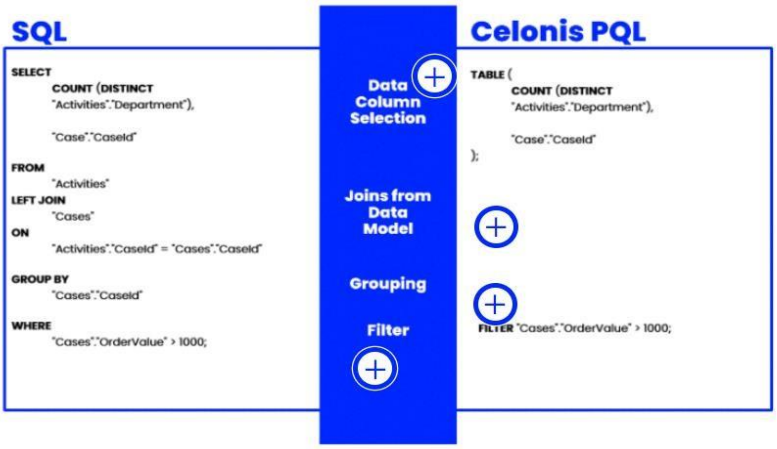

Fig No. 3.6: SQL VS PQL

3.3.3 Data into the EMS

In this topic we will study about two types they are

1. Set up a data pipeline
2. Refine your Data Pipeline

In the set up a data pipeline again divide into sub parts they are

1. Data Integration basics

2. Connect to Systems

3. Extract Data

4. Transform Data

5. Load a Data Model

In the Refine your Data Pipeline divide into parts they are

1.Schedule Data Jobs

2.Monitor and validate your Data pipeline

3.Multiple Process and Systems

4.Boost your EMS SQL Transformations

5.Connect Custom processes

6.Quality Assuring your Data Pipeline

1.Data Integration Basics

As a data engineer or analyst working in Data Integration (formerly known as Event Collection), you're responsible for bringing in clean, real-time process data into the EMS. In other words, you build the data pipeline.

2.Connect to systems

Connecting to source systems **is** your very first step to pull process data into the Celonis EMS. The EMS utilizes a broad set of technologies like message queues, Restful APIs, Soap APIs, direct database access, or system-specific solutions to connect.

Data Integration is where you set up connections and your data pipeline. The main ways you can bring data into the EMS are:

1. Process Connectors
2. Extractors (Data Connections)
3. Extractor Builder
4. File Uploads
5. Data Push API
6. Celoxtractor

3.Extract data

No matter which system you're working with when extracting data, it's a good idea to first understand the business process to know exactly which tables you need. We don't extract entire databases because

- take too long,
- be taxing on source systems,
- take up unnecessary cloud storage.

4.Transform Data

The Activity table represents your process and always contains at least these three columns that map your process:

- The object ID or case key,
- the process steps or activities that took place for the different case keys
- and the timestamps or event time of each activity

In the Purchase-to-Pay process, the Purchase Order Item Number is the central case key we follow. Every Purchase Order Item goes through different activities such as

creating the request, creating the item, receiving goods, and paying the invoice. And every activity has a corresponding event time.

In short, every Purchase Order Item has a unique case key that goes through various activities at different points in time. Together these three columns build the core of your process flow

CHAPTER 4
APPLICATIONS

Financial services, telecommunications, healthcare, and retail are just a few examples of industries where process mining can be used for business process management and process improvement. These sectors have a wealth of data that can be used as a starting point, and process deviations from their intended behavior can have expensive repercussions.

Financial Services: Because of the rise in transaction volume and the digitization of more industries, aberrant activity is harder to detect using manual methods. Companies in the financial services sector have the chance to continually and thoroughly identify issues within high-volume processes thanks to process mining, which is a solution to the increased regulatory and audit requirements.

Telecommunications: As subscriber quantities increase and activations become more and more automated, there is a greater danger of unsuccessful activations. When telecom companies get more orders, process mining gives them the chance to identify pricey issues and client blowback in their Order-to-Activation processes.

Healthcare: The risks associated with preserving population health and achieving individual patient journey objectives rise as data about patient experiences and results keep growing. Process mining supports the delivery of effective and high-quality end-to-end patient journeys for healthcare organizations dealing with the exponential growth of data, from before a first doctor appointment through treatment regimens to closed treatment cases.

Retail: Due to technology or process problems, retail businesses have seen expensive consumer fallout from complicated e-commerce operations. Process mining assists merchants in ensuring that consumers can complete transactions efficiently and without issues despite rising transaction volumes.

Digital Transformation: Process mining is frequently used in larger-scale digital transformation initiatives because it can give you the precise insights needed for process improvement, allowing systems to run more quickly, smoothly, and efficiently, as well as objective data-driven insights into the causes of delays and inefficiencies within business processes. As a result, process mining may assist in identifying the digital transformation opportunities with the greatest potential for value addition and determining whether or not transformation activities have really produced the desired results. To optimize returns on investments in projects for digital transformation, process mining becomes a crucial instrument.

Healthcare Management: In healthcare, process mining can be used to analyze patient pathways, identify treatment variations, and optimize hospital processes to enhance patient care and resource utilization.

Education: Process and task mining can reveal details on how users navigate on learning platforms to improve user experience for students. For example, process mining can show the potential root-causes behind students' exit rates from the given platform, such as the length of videos or organization of materials.

CHAPTER 5

REAL TIME APPLICATIONS AND USE CASE

The Celonis EMS capabilities are a combination of process intelligence and automation capabilities that help enterprises to find and capture business values by improving the performance of their core business processes. The EMS integrates in near real time with various sources with its built-in data integration capabilities. Celonis recently launched its advanced object centric process mining capability of Process sphere that captures interactions between multiple objects (eg., sales order and purchase order). The EMS also allows user to take targeted action through its action flow capabilities.

1.Accounts Payable Process: The AP process involves the management of invoices, payments, and supplier relationships. Process mining enables organizations to optimize KPIs by:

- **Streamlining Invoice Processing Time**: By analysing process flows, organizations can identify bottlenecks and high manual touchpoints where automation can help, reducing the time required to process invoices.
- **Improving Invoice Accuracy**: Process mining provides visibility into the invoice validation and approval process, allowing organizations to detect errors and implement corrective measures, leading to improved accuracy.
- **Enhancing On-time Payments**: Process mining helps identify delays in payment processing, quantify missed discounts and penalties and enable organizations to streamline workflows and ensure timely payments, improving supplier relationships.

Use case: One of the largest CPG firms deployed process mining and optimized the AP process by identifying payment delays and missed cash discounts resulting in $27M savings in working capital.

2.Accounts Receivables: Efficient management of the AR process is crucial for maintaining healthy cash flow. Process mining can enhance this process by uncovering inefficiencies, optimizing collections, and reducing payment delays from customers. Key KPIs for optimizing the Accounts Receivables process includes,

- **Reducing Days Sales Outstanding (DSO):** By analysing the collections process, process mining helps identify delays and bottlenecks, allowing organizations to streamline activities and shorten the collection cycle.
- **Enhancing Collection Efficiency**: Process mining provides insights into the collections process, enabling organizations to prioritize high-value customers and implement targeted strategies for improved collection effectiveness.
- **Accelerating Dispute Resolution**: Process mining helps identify common sources of disputes, enabling organizations to implement proactive measures and streamline dispute resolution processes, reducing resolution time.

Use case: A US based pharmaceutical company identified savings worth $29M in cash collection. They visualized the E2E accounts receivable process and identified gaps and inefficiencies that cause delays in collection.

3.IT Service Management (ITSM): The ITSM process involves the management of IT services, incidents, requests, and problem resolution. Process mining helps optimize KPIs within the ITSM process by:

- **Reducing Incident Resolution Time**: By analysing incident management workflows, process mining identifies bottlenecks and inefficiencies, allowing organizations to streamline incident resolution processes and shorten resolution times.
- **Improving First Call Resolution Rate**: Process mining provides insights into the ITSM process, enabling organizations to identify areas for improvement, enhance knowledge management systems.

Use case: A US based retailer leveraged process mining in their customer operations and optimized customer service, ITSM and onboarding process. This resulted in $30M /yearly savings by improving first time right KPI and enhanced the customer experience.

CHAPTER 6

LEARNING OUTCOMES

At the end of the course, one should be able to:

- Understand the fundamentals of process mining and its roleplay in discovering the bottlenecks and inefficiencies.
- Comprehend the various stages in drawing insights.
- Get familiar with process explorer and variate explorer
- Gain knowledge about PQL and PQL Engine architecture
- Know about the vast applications of Process Mining.
- Know about different process mining tools and the technical part of process mining

CONCLUSION

Process mining is a tool that businesses may rely on to conduct audits that are more accurate and efficient while avoiding conjecture and subjective conclusions. This results in less time being lost on doubt and subsequent retesting.

As technology continues to advance, process mining techniques are likely to become even more sophisticated and integrated with other data-driven approaches, further enhancing their ability to drive process excellence. However, successful implementation of process mining requires a comprehensive understanding of both the technology and the underlying business processes. Organizations that embrace process mining stand to gain a competitive edge by harnessing the power of data-driven insights to continuously refine their operations and achieve higher levels of efficiency and effectiveness

CERTIFICATE

REFERENCES:

1.Process Mining | Celonis | The Leader in Process Mining

2. AICTE Internship Enterprise Portal - Learning by doing. 1 Crore Internships by 2025 (aicte-india.org)

3. https://www.mdpi.com/1660-4601/16/7/1138

4.https://www.researchgate.net/publication/221470396_Process_Diagnostics_A_Method_Based_on_Process_Mining